NASA/TM-2003-

MODIS Validation, Data Merger and Other Activities Accomplished by the SIMBIOS Project: 2002-2003

Giulietta S. Fargion, Science Applications International Corporation, Maryland
Charles R. McClain, Goddard Space Flight Center, Greenbelt, Maryland

National Aeronautics and
Space Administration

Goddard Space Flight Center
Greenbelt, Maryland 20771

August 2003

Chapter 4

Diagnostic data set

Sean Bailey
FutureTech Corporation, Greenbelt, Maryland

At the first organizational meeting for the SIMBIOS program in 1995, a diagnostic data set for ocean color missions was conceived as a way to compare ocean color data across missions. The data set was to be created by each mission as part of routine processing and was to consist of spatial subsets with all relevant information necessary to produce derived products. These subsets were to be produced for a few selected sites. The diagnostic data set concept was revisited at several subsequent SIMBIOS science team meetings. At the third SIMBIOS science team meeting in September 1999, held in Annapolis, Maryland, the diagnostic data set concept took the first steps toward implementation with the selection of a number of proposed sites for the spatial subsets. The IOCCG working group on data merger met in January of 2000 and recommended a more complete list of sites for the data set. The list of sites was finalized at the fourth SIMBIOS Science Team meeting in January 2001.

Two conditions for the selection of a diagnostic data set site were formulated. First, a reliable source of *in situ* data (bio-optical and/or atmospheric) for the site must exist, and second, the principal investigator must be willing to share the *in situ* data with the SIMBIOS project. Sites used as vicarious calibration sources were ranked with the highest priority. Time series sites were ranked as priority 2. All other sites were ranked as priority 3. Several sites were recommended, but did not meet one or both of the defined criteria. Several of the sites were modified, either at the request of an investigator, in order to reduce redundancy or improve coverage (by reducing the amount of land included in the extracted data). The list of sites as currently implemented is found in Table 4.1 and Figure 4.1 shows a map of the locations.

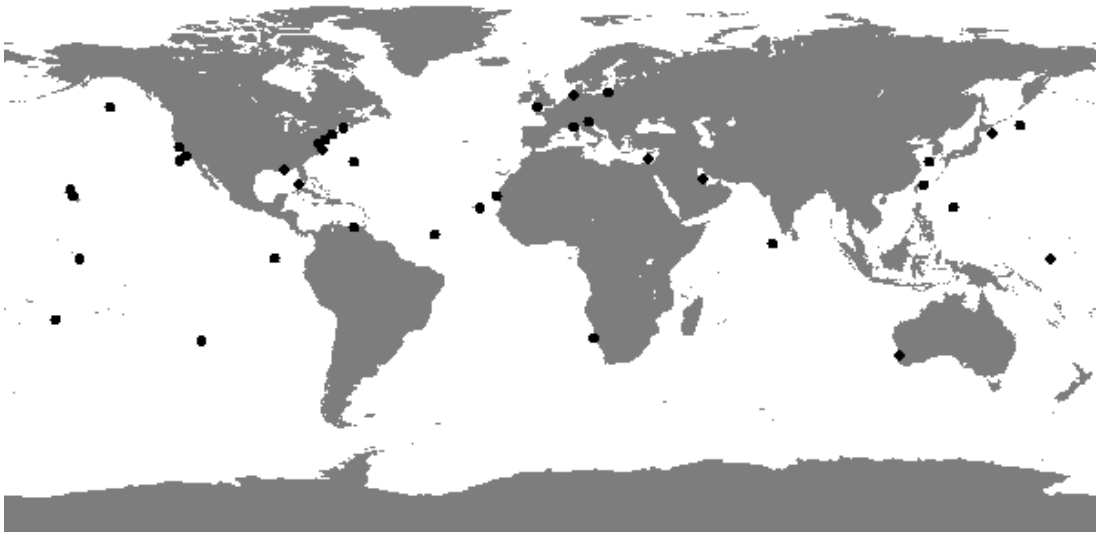

Figure 4.1: Diagnostic data set location.

By midyear 2001, the SIMBIOS project, in conjunction with the SeaWiFS project, had begun production of the L1A and L2 subsets of SeaWiFS LAC resolution data for the list of diagnostic data set sites. Prior to the start of the Collection 4 reprocessing of MODIS (Terra) data in March of 2002, the SIMBIOS project approached the MODIS Oceans Team with the request that MODIS produce a comparable set of extracted L1 and L2 data for inclusion in the diagnostic data set. The MODIS team agreed, however, as a consequence of the flow of data through MODAPS (the MODIS Data Production System), MODIS Oceans provides L1B and L2 extracts, rather than L1A as was the recommendation of the SIMBIOS Science Team and IOCCG working Group. The diagnostic data set files produced by MODAPS are sent to the SIMBIOS project for post processing. In order to ensure that only useful data are included in the diagnostic data set, a threshold on the number of valid pixels within the region of interest was set. If this threshold (currently 25%) is not met, the files are excluded from further processing. If the threshold is met, a L2 (chlorophyll) browse image and two TAR files are created. One TAR file for the L2 granules and one for the L1B granules. Since the MODIS data are produced in 5 minute granules, the region of interest for a given site may cross the boundary between two granules. When this occurs, all L2 products from both granules are placed in the same TAR file, likewise for the L1B granules. The TAR files are then compressed using gzip compression.